Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 9188–9193, August 1996
Microbiology

Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences

(archaebacteria/phylogeny/tree of life/molecular ecology)

SUSAN M. BARNS*, CHARLES F. DELWICHE†, JEFFREY D. PALMER‡, AND NORMAN R. PACE§

Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, IN 47405

Contributed by Norman R. Pace, April 24, 1996

ABSTRACT **Phylogenetic analysis of ribosomal RNA sequences obtained from uncultivated organisms of a hot spring in Yellowstone National Park reveals several novel groups of Archaea, many of which diverged from the crenarchaeal line of descent prior to previously characterized members of that kingdom. Universal phylogenetic trees constructed with the addition of these sequences indicate monophyly of Archaea, with modest bootstrap support. The data also show a specific relationship between low-temperature marine Archaea and some hot spring Archaea. Two of the environmental sequences are enigmatic: depending upon the data set and analytical method used, these sequences branch deeply within the Crenarchaeota, below the bifurcation between Crenarchaeota and Euryarchaeota, or even as the sister group to Eukaryotes. If additional data confirm either of the latter two placements, then the organisms represented by these ribosomal RNA sequences would merit recognition as a new kingdom, provisionally named "Korarchaeota."**

Phylogenetic analysis of molecular sequences has transformed our view of the course of evolution and allowed, for the first time, determination of a meaningful evolutionary history for prokaryotes (1). The most widely used molecule for phylogenetic analysis is small-subunit ribosomal RNA (rRNA). Most rRNA sequence diversity is found among microorganisms, both prokaryote and eukaryote; however, our understanding of microbial diversity has been incomplete because only a small fraction of naturally occurring microorganisms are routinely cultivatable in laboratory studies (2, 3). This technical hurdle has been overcome by the development of methods to obtain and analyze rRNA sequences from microbial communities *in situ*, without the requirement for laboratory cultivation of isolated species (4). The methods have allowed the identification and study of many previously undetected organisms (see ref. 1 for review).

Recent rRNA sequence-based analyses of the microbial constituents of a hydrothermal pool in Yellowstone National Park and of marine picoplankton have expanded the range of phenotypes and phylogenetic types of organisms known from the archaeal kingdom Crenarchaeota. Previously, this kingdom was thought to comprise an evolutionarily close-knit group of a few genera, united by an extremely thermophilic, sulfur-metabolizing phenotype (5). Discovery of rDNAs representing abundant, (presumably) low-temperature Crenarchaeota in temperate and Antarctic marine waters demonstrated greater physiological diversity than was previously known (6–9). Our initial study of the community of Obsidian Pool, a near-boiling (74–93°C), neutral pH hot spring (previously called "Jim's Black Pool") indicated the presence of at least 17 unknown organisms, most of which are only distantly related to cultivated crenarchaeal species (10).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

We report here the discovery of 19 additional novel archaeal sequences from this hot spring. These sequences, in combination with those reported previously, greatly expand the apparent diversity of Archaea. Among these sequences are two that group specifically with those of low-temperature marine Archaea, and two that branch prior to the diversification of all previously known Archaea in phylogenetic analyses that include both bacterial and eukaryotic outgroup taxa (but behave differently in analyses that exclude eukaryotes). This indicates that these sequences may represent a third, heretofore undescribed kingdom of Archaea.

MATERIALS AND METHODS

Ribosomal RNA gene (rDNA) sequences were obtained from organisms within the Obsidian Pool sediment community via polymerase chain reaction (PCR) amplification of total community DNA, cloning of products, and sequencing, essentially as described, using PCR primers 23FPL and 1391R (10). Secondary structure and computer analysis were used to screen for chimeric sequences (11, 12), and the single sequence identified as potentially artifactual by this method was excluded from subsequent analyses. Sequences obtained from rDNA clones were aligned manually with those obtained from the Ribosomal Database Project (RDP; ref. 13), R. Gutell (University of Colorado, Boulder) and E. DeLong (University of California, Santa Barbara), using rRNA secondary structure to identify homologous regions. Two alignments were assembled: one ("large"; 1620 positions) included all alignable positions, while the other ("small"; 922 positions) included only those sites alignable among all domains. (All rRNA sequence alignments used in this analysis are available at http://www.bio.indiana.edu/~nrpace/pacelab/published/and from the authors.) Distance matrix analysis of sequences was performed by using PHYLIP (version 3.5, distributed by J. Felsenstein, University of Washington, Seattle). Distance matrices were calculated with the program DNADIST by using the correction for multiple superimposed mutations and site-to-site rate variation of Jin and Nei (14) with a shape parameter "a" of 1.3, determined by maximum likelihood estimation with PAUP version 4.0d40. Trees based on distance

Abbreviations: ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor joining distance matrix; rDNA, rRNA-encoding genes; RDP, Ribosomal Database Project.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. 463338–463371).

*Present address: Environmental Molecular Biology, M888 Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545. e-mail: sbarns@telomere.lanl.gov.

†Present address: Department of Plant Biology, H. J. Patterson Hall, University of Maryland, College Park, MD 70742-5815. e-mail: delwiche@bio.indiana.edu.

‡e-mail: jpalmer@bio.indiana.edu.

§To whom reprint requests should be sent at the present address: Department of Plant and Microbial Biology, Koshland Hall 111, University of California, Berkeley, CA 94720. e-mail: nrpace@nature.berkeley.edu.

data were inferred by the neighbor-joining (NEIGHBOR) method (NJ), and bootstrap analyses utilized 300 replicate data sets. Unweighted maximum parsimony (MP) analyses were performed using PAUP (versions 3.1.1 and 4.0d29; D. L. Swofford, Smithsonian Institution), with trees found by 100 replicates of random-sequence addition heuristic searches and tree-bisection-reconstruction (TBR) branch swapping on shortest trees. Maximum likelihood (ML) analyses were performed by using FASTDNAML (version 1.0, distributed by RDP; refs. 15 and 16), employing empirical base frequencies. The optimal transition/transversion ratio (T = 1.2) was estimated by comparing the likelihood under a T of 1.0, 1.1, ... 2.0 for an initial tree calculated with T = 2.0. Nonbootstrap trees were optimized by global rearrangements of branches. Bootstrapping under ML and MP was performed with 100 replicates using a random addition sequence. An auxiliary program, DNARATES (13), was used to calculate site-specific rates of nucleotide substitution for each position in the alignment, and these rates were used in some ML analyses and in determining rate categories for analyses that excluded rapidly evolving positions. Since sequences of widely varying base composition (41–68% G+C) were analyzed, NJ and MP analysis of transversion events alone (17), and NJ analysis of distance matrices produced by LOGDET (18) using all characters and with a correction for the estimated fraction of invariant positions, also were used. The fraction of invariant sites was estimated by ML estimation with PAUP version 4.0.d34 using the F84 model, empirical frequencies, and T = 1.2.

RESULTS AND DISCUSSION

Phylogenetic Analysis of Obsidian Pool rRNA Sequences. The abundance of previously unknown archaeal sequences detected in the first analysis of the Obsidian Pool community (10) indicated that further survey likely would identify additional novel types. In the present analysis, 141 insert-containing clones were screened by single-nucleotide (ddG) sequencing, and 26 distinct sequence types were identified. Only five of these (pJP6, pJP8, pJP27, pJP74, and pJP78) were among the 11 sequences analyzed in the first study. Approximately 500 nt of sequence were obtained from each of the 21 new rDNA types, and these were used in preliminary phylogenetic analyses (not shown) together with sequences from the first analysis and from the RDP data base. Thirteen Obsidian Pool clones representative of distinct evolutionary groups were sequenced (≈1330 nt). A phylogenetic tree of all full-length and partial sequences from both Obsidian Pool analyses (32 sequences total) was inferred by ML analysis (Fig. 1).

Most of the sequences obtained from Obsidian Pool affiliated with the crenarchaeal kingdom of Archaea (Fig. 1). A few of the sequences (pSL91, pJP74, pJP7, pJP8, pSL60, pJP6, pJP81, pJP9, and pSL23) were highly similar to the rRNA sequences of cultivated Archaea, but none was identical to any available sequence from a cultivated organism. Representatives of all previously known families of Crenarchaeota are present in this hot spring. Two sequences (pSL12 and pSL77) group specifically with SBAR5, a sequence from recently discovered Crenarchaeota of pelagic marine picoplankton (ref. 6; see below). Only two sequences of euryarchaeal affiliation were obtained (pJP9 and pSL23). These were highly similar to that of the thermophilic sulfate-reducing marine organism *Archaeoglobus fulgidus* (19). The majority of the sequences, however, show no close affiliation with the rRNA of any cultured Archaea. Most of these sequences branch more deeply from the crenarchaeal line of descent than do those of cultured species. The kingdom Crenarchaeota had been thought to be restricted to a few genera, closely related in physiology and phylogeny (5). Consequently, these novel lineages substantially expand the known phylogenetic diversity of Crenarchaeota.

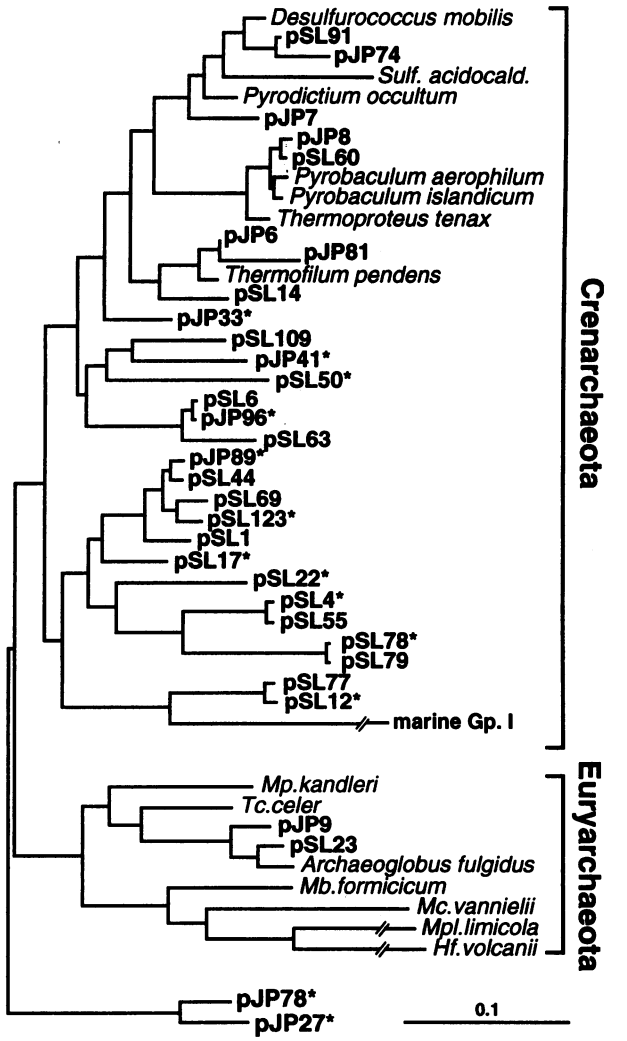


FIG. 1. Phylogenetic tree of rDNA clones from Obsidian Pool, illustrating the wide diversity of Archaea present in the community. Sequences recovered from sediment organisms (designated "pJP" and "pSL") were analyzed with sequences from cultured archaeal species (in italics) obtained from the RDP data base. "Marine SBAR5" sequence is that of clone SBAR5 recovered from Pacific marine bacterioplankton by DeLong (6). Tree was inferred by ML analysis as described in the text, and is arbitrarily rooted in the pJP27/pJP78 lineage. Scale bar represents 10 mutations per 100-nt sequence positions. Sequences of clones designated with asterisks (*) were determined in their entirety, and eight of these were used in the three domain analysis of Fig. 2. *Sulf. acidocald.*, *Sulfolobus acidocaldarius*; *Mb. formicicum*, *Methanobacteriam formicicum*; *Mc. vannielii*, *Methanococcus vannielii*; *Mpl. limicola*, *Methanoplanus limicola*; *Hf. volcanii*, *Haloferax volcanii*.

Three-Domain Analyses of Obsidian Pool Sequences. To examine the evolutionary relationships of the Obsidian Pool sequences to members of the domains Archaea, Bacteria, and Eucarya, to provide adequate outgroup sequences, and to determine the effect of inclusion of these new, deeply branching sequences on the topology of the universal tree, eight representative Obsidian Pool sequences (pSL4, pSL12, pSL17, pSL22, pSL50, pJP27, pJP78, and pJP96) were aligned with those of 56 additional species representing the three domains. Taxa were chosen to sample broadly the diversity within each domain and to emphasize deep-branching sequences. The ML tree determined from the 1620-nt data set is shown in Fig. 2. Since phylogenetic analysis can be affected by the assumptions inherent in the analytical technique chosen (22), we compared trees constructed with NJ, MP, and ML algorithms, and

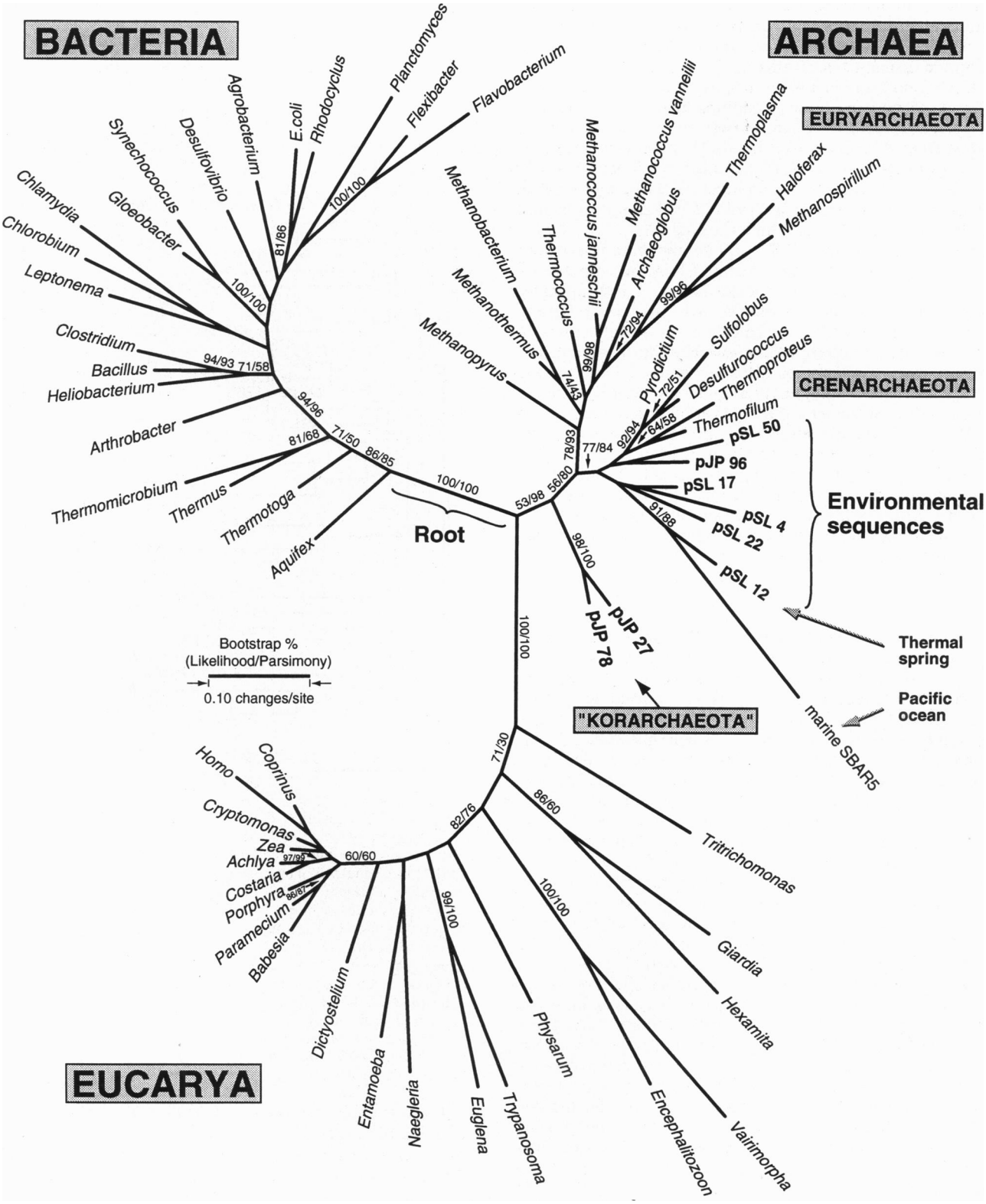


FIG. 2. "Universal" unrooted phylogenetic tree showing positions of major groups of sequences recovered from the Obsidian Pool community. Tree was inferred by ML analysis of 1620 homologous nucleotide positions of sequence from each organism or clone. Numbers indicate percentage of bootstrap resamplings that support indicated branches in ML (before slash) and MP (after slash) analyses for those groups only that attained >60% support with at least one of the two methods. Analyses of duplicated protein genes have placed the root of the tree on the branch at the base of the Bacteria (20, 21).

bootstrap analysis was performed wherever practical (23). Because of concern that site-to-site rate variation might affect the analysis, Jin and Nei correction (14) was used to calculate distances for the distance analyses presented here, and DNARATES was employed to determine rate categories for some ML analyses (13). As expected, most conflicts between trees determined with different analytical methods occurred in portions of the tree having weak bootstrap support.

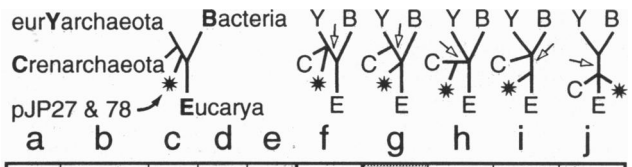
Within the bacterial domain, thermophilic species were found as the most deeply branching sequences, although

bootstrap support for this was lower in the four analyses which compensate for base-compositional bias (data not shown). *Aquifex* is supported as the most basal lineage (24, 25), and a few bacterial groups were strongly supported (Fig. 2). However, relative branching orders among most of the bacterial groups were poorly resolved, consistent with previous analyses using rRNA (26, 27) and other molecules (28–30). Among Eucarya, sequences from organisms lacking mitochondria (*Tritrichomonas*, the diplomonads *Giardia* and *Hexamita*, and the microsporidia *Vairimorpha* and *Encephalitozoon*) branch most deeply. Consistent with other small subunit rRNA analyses (31), however, the branching order among these taxa was not strongly supported. Other features of the eukaryotic topology also are generally consistent with previous rRNA analyses (31), although many of these features do not have strong bootstrap support.

Analyses using both the 1620-character and the 922-character data sets yielded generally similar results. In the archaeal domain, the larger character set showed enhanced resolution among the terminal taxa, but reduced resolution at the base of the domain (Fig. 3). In analyses utilizing either data set, the tree was fairly well resolved within the archaeal domain. The clades of *Thermoproteus/Thermofilum* and *Pyrodictium/Desulfurococcus/Sulfolobus* are consistent with previous analyses (32), as is the association between the phenotypically disparate Methanomicrobiales *(Methanospirillum)* and extreme halophiles *(Haloferax)* (33). The gene *rpoB* is divided into two parts in *Archaeoglobus*, halophiles and methanogens (although *rpoB* in *Methanopyrus* has not been studied), implying that these taxa form a group to the exclusion of *Thermococcus, Thermoplasma,* and all other taxa (34), but we did not observe such a group in our analyses. *Methanopyrus* consistently appears as the first branch in the Euryarchaeota, as previously seen (35). Six of the eight Obsidian Pool sequences analyzed (pSL4, pSL12, pSL17, pSL22, pSL50, and pJP96) affiliate consistently with the Crenarchaeota, but branch more deeply from the crenarchaeal lineage than any cultivated species.

Relationship of Obsidian Pool Sequences to Marine Crenarchaeota. This study also identified close, probably thermophilic relatives (pSL12 and pSL77) of recently discovered, abundant Crenarchaeota from diverse coastal and open ocean sites in the North Pacific and Antarctic (represented by sequence "marine SBAR5" in Figs. 1 and 2; refs. 6–8). Previous analyses were unable to resolve the phylogenetic position of these marine rRNA sequences within the Crenarchaeota, but indicated that they constitute a deeply branching group (6, 9). In the present analysis, the Obsidian Pool pSL12 sequence always formed a well-supported clade with the marine sequence (Fig. 2). Although the marine sequence has a lower G+C content (51% G+C in the large character set) than those of other Crenarchaeota (58–68% G+C), this was found to have little effect on the strength of the grouping in transversion and LOGDET analyses (not shown). The relatively low G+C composition of this and other marine archaeal rRNA sequences, together with their cold-water source, indicate that the organisms contributing these sequences, in remarkable contrast to all other characterized Crenarchaeota, are probably not thermophilic. The higher G+C content of the pSL12 sequence (58%), as well as its hot spring source, indicate origin from a thermophilic organism. The specific affiliation of the pSL12 and marine sequences and their nested position within other thermophilic lineages implies that the low-temperature marine Archaea are descendants of ancestral thermophiles. The uniform occurrence of thermophilic lineages near the base of the archaeal tree indicates that the common ancestor of the Archaea was thermophilic and is consistent with the theory of a high-temperature origin of life (1).

Analysis of Archaeal Monophyly. The three-domain alignments also were used to examine the question of monophyly of Archaea in light of the extensive phylogenetic diversity intro-



a	b	c	d	e	f	g	h	i	j
64	small		ML		74	70	13	13	11
64	small		ML	DR	-	-	-	-	-
64	small		MP		99	93			
64	small	tv	MP		95	52	18		
64	small		NJ	JN	50	69		48	26
64	small	tv	NJ	JN	76	59	15	11	17
64	small		NJ	LD	95	92			
64	small	ni	NJ	LD	72	91			
64	large		ML		53	56		38	28
64	large		ML	DR	-	-	-	-	-
64	large		MP		98	80	18		
64	large		NJ	JN	38	16	59	20	14
45	large		ML		57	54	20	32	21
45	large		ML	DR	-	-	-	-	-
45	large	1-4	ML		60	64		28	
45	large	1-3	ML		74	74		24	
45	large		MP		98	92			
43†	large		ML		72				
43†	large		MP		99				
AE	small		ML			-	-		
AE	small		ML	DR		-	-		
AE	small		MP			70	20		
AE	small		NJ	JN		95			
AB	small		ML			47	48		
AB	small		ML	DR		-	-		
AB	small	1-4	ML			58	16		
AB	small	1-3	ML			72	12		
AB	small		MP			70	28		
AB	small		NJ	JN		39	61		
AB	large		ML			33	63		
AB	large		ML	DR		-	-		
AB	large	1-4	ML			18	64		
AB	large	1-3	ML			21	29		
AB	large		MP			74	25		
AB	large		NJ	JN			84		

FIG. 3. Summary of phylogenetic analyses, emphasizing position of pJP27/pJP78 lineage. Key features of the trees found by different phylogenetic methods and with different taxon and character sets are shown. Each column (f–j) refers to the branch indicated by the arrow on the tree at the top of the column. The topology of the best tree(s) found is indicated by shading (except in column f). C, Crenarchaeota; Y, Euryarchaeota; *, pJP27/pJP78; E, Eucarya; B, Bacteria. Numbers indicate bootstrap support for that topology, blank boxes indicate a bootstrap value below 10%, and dashes indicated analyses where bootstrap analysis was not performed. The columns are as follows. (a) Number of sequences used in analysis. AB, archaeal and bacterial sequences only; AE, archaeal and eukaryotic sequences only; †, sequences pJP27 and pJP78 excluded from analysis. (b) Size of data set used in analysis. Small, 922 unambiguously alignable nucleotides for each sequence; large, 1620 positions (1394, 1351, and 1098 nt of sequence for archaeal, bacterial, and eukaryotic taxa, respectively). (c) Character selection. tv, Analysis of transversion events only; ni, no invariant positions analyzed; 1–4 and 1–3, analysis of rate categories 1–4 and 1–3, respectively, as determined by DNARATES. (d) Analysis method. ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor joining distance matrix. (e) Method of data correction. DR, maximum likelihood site-to-site rate correction by DNARATES; JN, distance matrix site-to-site rate correction; LD, LOGDET correction for base compositional bias. (f) Bootstrap support for monophyly of Archaea, including pJP27 and pJP78 sequences. The best tree is not shown because this information is given in columns g and h. (g) Branching of pJP27/pJP78 clade from below crenarchaeal/euryarchaeal split. (h) pJP27/pJP78 grouping with Crenarchaeota. (i) pJP27/pJP78 lineage grouping with Eucarya. (j) pJP27/pJP78 grouping with Crenarchaeota and Eucarya. Values below 50% are considered unreliable.

duced by the Obsidian Pool sequences. Previous analyses of rRNA data generally have been limited both in use of consistency measures (i.e., bootstrap analysis) and/or number of archaeal sequences analyzed. Most analyses of rRNA (and other) sequences have indicated that Archaea constitute a monophyletic group (20, 27, 36). However, some analyses of protein sequences (21) and of rRNA sequences using different techniques (37) group Crenarchaeota with Eucarya, to the exclusion of Euryarchaeota and Bacteria. Only moderate (52–78%) bootstrap support for archaeal monophyly (with respect to Eucarya) has been reported recently for large subunit rRNA data (27), while the latest elongation factor analyses modestly support (35–89%) archaeal paraphyly (21). Whereas the bacterial and eukaryotic domains were monophyletic with 100% bootstrap support in all of our analyses, support for monophyly of Archaea varied with the analytical method used (Fig. 3, column f). All parsimony analyses strongly support archaeal monophyly (95–99%), while support for monophyly is lower in the ML (53–82%) and NJ analyses (38–89%). Similar results were seen when pJP27 and pJP78 were excluded from analyses (Fig. 3). Nonetheless, there is little or no bootstrap support (≤28%; Fig. 3, column j) for the association of Crenarchaeota with eukaryotes, to the exclusion of Euryarchaeota and Bacteria (the "eocyte tree"; ref. 37).

The issue of archaeal monophyly is best viewed as a problem in the placement of eukaryotes with respect to Archaea. Because of the very long branch leading to the eukaryotes and the extreme rRNA divergence within this domain, it is to be expected that this group will be difficult to place with confidence (29, 38). The difference between archaeal monophyly and paraphyly involves movement of the eukaryotic branch across a single node involving two long branches (leading to Eucarya and Bacteria) and two short branches (leading to Crenarchaeota and Euryarchaeota). Thus, although small subunit rRNA data favor archaeal monophyly, unequivocal resolution of the origin of eukaryotes will require further data from additional genes. This analysis also highlights the need for better molecular phylogenetic methods for resolving relationships among anciently diverging lineages.

Phylogenetic Position of Sequences pJP27 and pJP78. The most problematic sequences in our analysis were those from pJP27 and pJP78, which consistently group together. Most of the phylogenetic analyses carried out in the initial report (10) on Obsidian Pool placed the pJP27 and pJP78 sequences in the Crenarchaeota, but bootstrap support was modest in the two analyses presented (57% and 76%). The few analyses that placed them below the Crenarchaeota/Euryarchaeota split had low (<50%) support. Hence, the pJP27 and pJP78 sequences have been regarded as members, albeit deeply branching ones, of the Crenarchaeota (10, 39). The larger and more thorough analyses reported here indicate a more complex situation.

The phylogenetic position of pJP27 and pJP78 varies depending upon the data set and analytical method used. In analyses that include eukaryotes, these taxa generally fall on the archaeal lineage, but below the bifurcation between Crenarchaeota and Euryarchaeota. This topology occurred, with moderate to high bootstrap support, in the majority of analyses including eukaryotes, but a few such analyses showed pJP27 and pJP78 branching with Eukarya, albeit with little bootstrap support. By contrast, in analyses that exclude eukaryotes, the pJP27/pJP78 lineage branches below the Crenarchaeota/Euryarchaeota split in some cases, but groups with the Crenarchaeota in others (Fig. 3, columns g and h). Because of the very long eukaryotic branches, it is not clear whether inclusion of eukaryotes adds phylogenetic information to the analysis, allowing clarification of an otherwise unresolved polytomy, or adds noise, resulting in a long-branch artifact (38). Analyses that exclude eukaryotes, however, are inherently unable to address their position with respect to pJP27 and pJP78.

Since the generally more conservative characters in the 922-character data set more strongly supported the deep placement of pJP27/pJP78, we examined the effects of site-to-site rate variation. Conservative characters as determined by DNARATES, however, did not conclusively support any one topology (Fig. 3). Thus three placements of the pJP27/pJP78 lineage must be considered: within the Crenarchaeota, below the bifurcation between Crenarchaeota and Euryarchaeota, or with the Eukarya. Ironically, the topology modestly favored by elongation factor data, with Eukarya sister taxon to the Crenarchaeota in a paraphyletic Archaea (21, 40), although seen only rarely in our analyses, is consistent with the two extreme placements of pJP27 and pJP78 in these rRNA analyses, within the Crenarchaeota and as sister taxon to the Eukarya.

Although the present analyses do not unequivocally determine the placement of the pJP27/pJP78 lineage, they clearly indicate that this is a noteworthy group. While it is possible that they are simply another long-branched crenarchaeal group (analogous to the previously unplaced marine taxa), the data raise the strong possibility that these sequences represent an uncharacterized fundamental lineage of Archaea (or perhaps Eucarya). Because rRNA data favor placement of the pJP27 and pJP78 among a monophyletic archaea, we propose to recognize this lineage provisionally as a third archaeal kingdom, the *Korarchaeota* [Greek noun κορος (koros) "young man" or κορη (kore) "young woman," for the early divergence of this group during the evolution of Archaea; Greek adjective αρχαιος (archaios) "ancient, primitive"]. Acceptance or rejection of this status and name will require phylogenetic analysis of other gene sequences from this lineage. These sequences, as well as an understanding of the physiology and other properties of the organisms represented by the pJP27 and pJP78 sequences, will be obtained most readily once these creatures are cultivated (efforts are underway; S. Burggraf, personal communication). Even if cultivation proves intractable, methods of large-fragment cloning should allow access to the genomes of the organisms (41). Recently, a third member of this group was detected in a separate hot spring environment (A.-L. Reysenbach, personal communication), suggesting that these organisms are widespread in thermal habitats. A greater understanding of them may provide an additional perspective on the nature of life at high temperatures, today and at the time of the origin of all life.

We gratefully acknowledge Ed DeLong, Mitchell Sogin, and Robin Gutell for providing aligned sequence data; Carl Woese, Gary Olsen, Joe Felsenstein, Pete Lockhart, Sandie Baldauf, Ford Doolittle, Michael Gray, David Spencer, Mitchell Sogin, Charles Marshall, Séan Turner, Walter Fitch, and David Hillis for extensive helpful comments on phylogenetic analyses; and Carl Woese, Gary Olsen, Dan Frank, Ed DeLong, François Lutzoni, Kathleen Pryer, and Peter Kuhlman for comments on the manuscript. We also thank Bernadette Pace for supplying *Thermus aquaticus* DNA polymerase, Matthew Jeffries for sequencing and technical assistance, Bill Hansen for help with Greek nomenclature, Steve Koch and Anna-Louise Reysenbach for field assistance, and the staff of Yellowstone National Park, especially Bob Lindstrom, for their enthusiastic cooperation. Indiana University's Center for Innovative Computer Applications provided computer time and technical support. This study was supported by a National Science Foundation Graduate Fellowship to S.M.B., U.S. Department of Energy Grant DE-FG02-93ER-20088 and National Institutes of Health Grant GM34527 to N.R.P., and National Institutes of Health Grant GM-35087 to J.D.P.

1. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
2. Ward, D. M., Bateson, M. M., Weller, R. & Ruff-Roberts, A. L. (1992) *Adv. Microb. Ecol.* **12,** 219–286.
3. Amann, R. I., Ludwig, W. & Schleifer, K. (1995) *Microbiol. Rev.* **59,** 143–169.
4. Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J. (1986) *Adv. Microb. Ecol.* **9,** 1–55.

5. Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci USA* **87,** 4576–4579.
6. DeLong, E. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5685–5689.
7. Fuhrman, J. A., McCallum, K. & Davis, A. A. (1992) *Nature (London)* **356,** 148–149.
8. DeLong, E. F., Wu, K. Y., Prezelin, B. B. & Jovine, R. V. M. (1994) *Nature (London)* **371,** 695–697.
9. McInerney, J. O., Wilkinson, M., Patching, J. W., Embley, T. M. & Powell, R. (1995) *Appl. Environ. Microbiol.* **61,** 1646–1648.
10. Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1609–1613.
11. Liesack, W., Weyland, H. & Stackebrandt, E. (1991) *Microb. Ecol.* **21,** 191–198.
12. Larsen, N., Olsen, G. J., Maidak, B. L., McCaughey, M. J., Overbeek, R., Macke, T. J., Marsh, T. L. & Woese, C. R. (1993) *Nucleic Acids Res.* **21,** 3021–3023.
13. Maidak, B. L., Larsen, N., McCaughey, M. J., Overbeek, R., Olsen, G. J., Fogel, K., Blandy, J. & Woese, C. R. (1994) *Nucleic Acids Res.* **22,** 3485–3487.
14. Jin, L. & Nei, M. (1990) *Mol. Biol. Evol.* **7,** 82–103.
15. Olsen, G. J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994) *Comput. Appl. Biosci.* **10,** 41–48.
16. Felsenstein, J. (1981) *J. Mol. Evol.* **17,** 368–376.
17. Woese, C. R., Achenbach, L., Rouviere, P. & Mandelco, L. (1991) *Syst. Appl. Microbiol.* **14,** 364–371.
18. Lockhart, P., Steel, M. A., Hendy, M. D. & Penny, D. (1994) *Mol. Biol. Evol.* **11,** 605–612.
19. Burggraf, S., Jannasch, H. W., Nicolaus, B. & Stetter, K. O. (1990) *Syst. Appl. Microbiol.* **13,** 24–28.
20. Brown, J. R. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2441–2445.
21. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7749–7754.
22. Swofford, D. L. & Olsen, G. J. (1991) in *Molecular Systematics,* eds. Hillis, D. & Moritz, C. (Sinauer, Sunderland, MA), pp. 411–510.
23. Felsenstein, J. (1988) *Annu. Rev. Genet.* **22,** 521–565.
24. Burggraf, S., Olsen, G. J., Stetter, K. O. & Woese, C. R. (1992) *Syst. Appl. Microbiol.* **15,** 352–356.
25. Reysenbach, A.-L., Wickham, G. S. & Pace, N. R. (1994) *Appl. Environ. Microbiol.* **60,** 2113–2119.
26. Peer, Y. V. d., Neefs, J.-M., Rijk, P. D., Vos, P. D. & Wachter, R. D. (1994) *Syst. Appl. Microbiol.* **17,** 32–38.
27. Rijk, P. D., Peer, Y. V. d., Broeck, I. V. d. & Wachter, R. D. (1995) *J. Mol. Evol.* **41,** 366–375.
28. Delwiche, C. F., Kuhsel, M. & Palmer, J. D. (1995) *Mol. Phylogenet. Evol.* **4,** 110–128.
29. Viale, A. M., Arakaki, A. K., Soncini, F. C. & Ferreyra, R. G. (1994) *Int. J. Syst. Bacteriol.* **44,** 527–533.
30. Lloyd, A. T. & Sharp, P. M. (1993) *J. Mol. Evol.* **37,** 399–407.
31. Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. (1993) *Mol. Biochem. Parisitol.* **59,** 41–48.
32. Olsen, G. J. & Woese, C. W. (1995) *FASEB J.* **7,** 113–123.
33. Burggraf, S., Stetter, K. O., Rouviere, P. & Woese, C. R. (1991) *Syst. Appl. Microbiol.* **14,** 346–351.
34. Woese, C. R. (1993) in *Biochemistry of Archaea (Archaebacteria),* eds. Kates, M., Kushner, D. J. & Matheson, A. T. (Elsevier, Amsterdam), pp. vii–xlviii.
35. Burggraf, S., Stetter, K. O., Rouviere, P. & Woese, C. R. (1991) *Syst. Appl. Microbiol.* **14,** 346–351.
36. Olsen, G. J. & Woese, C. W. (1989) *Can. J. Microbiol.* **35,** 119–123.
37. Lake, J. A. (1991) *Trends Biochem. Sci.* **16,** 46–49.
38. Felsenstein, J. (1978) *Syst. Zool.* **27,** 401–410.
39. Olsen, G. J. (1994) *Nature (London)* **371,** 657–658.
40. Rivera, M. C. & Lake, J. A. (1992) *Science* **257,** 74–76.
41. Stein, J. L., Marsh, T. L., Wu, K.-Y., Shizua, H. & DeLong, E. F. (1996) *J. Bacteriol.* **178,** 591–599.